RWTH
LTCS–Report

Aachen University of Technology
Research group for
Theoretical Computer Science

Computing Most Specific Concepts in Description Logics with Existential Restrictions

Ralf Küsters and Ralf Molitor

LTCS-Report 00-05

RWTH Aachen
LuFg Theoretische Informatik
http://www-lti.informatik.rwth-aachen.de

Ahornstr. 55
52074 Aachen
Germany

Computing Most Specific Concepts in Description Logics with Existential Restrictions

Ralf Küsters
Institut für Informatik und Praktische Mathematik
Christian-Albrechts-Universität zu Kiel
kuesters@ti.informatik.uni-kiel.de

Ralf Molitor
IT Research and Development Group
Swiss Life
Zurich, Switzerland
ralf.molitor@swisslife.ch

Abstract

Computing the most specific concept (msc) is an inference task that can be used to support the "bottom-up" construction of knowledge bases for KR systems based on description logics. For description logics that allow for number restrictions or existential restrictions, the msc need not exist, though. Previous work on this problem has concentrated on description logics that allow for universal value restrictions and number restrictions, but not for existential restrictions. The main new contribution of this paper is the treatment of description logics with existential restrictions. More precisely, we show that, for the description logic $\mathcal{ALE}$ (which allows for conjunction, universal value restrictions, existential restrictions, negation of atomic concepts, as well as the top and the bottom concept), and its sublanguages $\mathcal{EL}$ (which allows for conjunction, existential restrictions and the top-concept) and $\mathcal{EL}_\neg$ (which extends $\mathcal{EL}$ by negation of atomic concepts) the msc of an ABox-individual only exists in case of acyclic ABoxes. For cyclic ABoxes, we show how to compute an approximation of the msc.

Our approach for computing the (approximation of the) msc is based on representing concept descriptions by certain trees and ABoxes by certain graphs, and then characterizing instance relationships by homomorphisms from trees into graphs. The msc/approximation operation then mainly corresponds to unraveling the graphs into trees and translating them back into concept descriptions.

1 Introduction

The most specific concept (msc) of an individual b is a concept description that has b as instance and is the least concept description (w.r.t. subsumption) with this property. Roughly speaking, the msc is the concept description that, among all concept descriptions of a given DL, represents b best. Closely related to the msc is the least common subsumer (lcs), which, given concept descriptions $C_1, \ldots, C_n$, is the least concept description (w.r.t. subsumption) subsuming $C_1, \ldots, C_n$. Thus, where the msc generalizes an individual, the lcs generalizes a set of concept descriptions.

In [2, 3, 4], the msc (first introduced in [15]) and the lcs (first introduced in [5]) have been proposed to support the bottom-up construction of a knowledge base. The motivation comes from an application in chemical process engineering [17], where the process engineers construct the knowledge base (which consists of descriptions of standard building blocks of process models) as follows: First, they introduce several "typical" examples of a standard building block as individuals, and then they generalize (the descriptions of) these individuals into a concept description that (i) has all the individuals as instances, and (ii) is the most specific description satisfying property (i). The task of computing concept descriptions satisfying (i) and (ii) can be split into two subtasks: computing the msc of a single individual, and computing the lcs of a given finite number of concepts.

The lcs has been thoroughly investigated for (sublanguages of) CLASSIC [5, 2, 13, 11], for DLs allowing for existential restrictions like $\mathcal{ALE}$ [3], and most recently, for $\mathcal{ALEN}$, a DL allowing for both existential and number restrictions [14]. For all these DLs, except for CLASSIC in case attributes are interpreted as total functions [13], it has turned out that the lcs always exists and that it can effectively be computed. Prototypical implementations show that the lcs algorithms behave quite well in practice [7, 4].

For the msc, the situation is not that rosy. For DLs allowing for number restrictions or existential restrictions, the msc does not exist in general. Hence, the first step in the bottom-up construction, namely computing the msc, cannot be performed. In [2], it has been shown that for $\mathcal{ALN}$, a sublanguage of CLASSIC, the existence of the msc can be guaranteed if one allows for cyclic concept descriptions, i.e., concepts with cyclic definitions, interpreted by the greatest fixed-point semantics. Most likely, such concept descriptions would also guarantee the existence of the msc in DLs with existential restrictions. However, current DL-systems, like FaCT [10] and RACE [9], do not support cyclic concept descriptions; although they allow for cyclic definitions of concepts, these systems do not employ the greatest fixed-point semantics, but only descriptive semantics. Consequently, cyclic concept descriptions returned by algorithms computing the msc cannot be processed by these systems.

In this paper, we therefore propose to approximate the msc. Roughly speaking, for some given non-negative integer k, the *k-approximation* of the msc of an individual b is the least concept description (w.r.t. subsumption) among all concept descriptions with b as instance and role depth at most k. That is, the set of potential msc's is restricted to the set of concept descriptions with role depth bounded by k. For (sublanguages of) $\mathcal{ALE}$ we show that k-approximations always exist and that they can effectively be computed. Thus, when replacing "msc" by "k-approximation", the first step of the bottom-up construction can always be carried out. Although the original outcome of this step is only approximated, this might in fact suffice as a first suggestion to the knowledge engineer.

While for full $\mathcal{ALE}$ our k-approximation algorithm is of questionable practical use (since it employs a simple enumeration argument), we propose improved algorithms for the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$ of $\mathcal{ALE}$. ($\mathcal{EL}$ allows for conjunction and existential restrictions, and $\mathcal{EL}_\neg$ additionally allows for a restricted form of negation.) Our approach for computing k-approximations in these sublanguages is based on representing concept descriptions by certain trees and ABoxes by certain (systems of) graphs, and then characterizing instance relationships by homomorphisms from trees into graphs. The k-approximation operation then consists in unraveling the graphs into trees and translating them back into concept descriptions. In case the unraveling yields finite trees, the corresponding concept descriptions are "exact" msc's, showing that in this case the msc exists. Otherwise, pruning the infinite trees on level k yields k-approximations of the msc's.

The outline of the paper is as follows. In the next section, we introduce the basic notions and formally define k-approximations. To get started, in Section 3 we present the characterization of instance relationships in $\mathcal{EL}$ and show how this can be employed to compute k-approximations or the msc (if it exists). In the subsequent section we extend the results to $\mathcal{EL}_\neg$, and finally deal with $\mathcal{ALE}$ in Section 5. The paper concludes with some remarks on future work.

2 Preliminaries and known results

Concept descriptions are inductively defined with the help of a set of *constructors*, starting with a set N_C of *concept names* and a set N_R of *role names*. The constructors determine the expressive power of the DL. In this work, we consider concept descriptions built from the constructors shown in Table 1. In the description logic $\mathcal{EL}$, concept descriptions are formed using the constructors top-concept ($\top$), conjunction ($C \sqcap D$) and existential restriction ($\exists r.C$). The description logic $\mathcal{EL}_\neg$ additionally provides us with primitive negation ($\neg P$, $P \in N_C$), and $\mathcal{ALE}$ allows for all the constructors shown in Table 1 except of number restrictions. Finally, $\mathcal{ALN}$ allows for the top- and bottom-concept, concept conjunction, primitive negation, value restrictions, and number restrictions.

The semantics of a concept description is defined in terms of an *interpretation* $\mathcal{I} = (\Delta, \cdot^\mathcal{I})$. The domain Δ of $\mathcal{I}$ is a non-empty set of individuals and the interpretation function $\cdot^\mathcal{I}$ maps each concept name $P \in N_C$ to a set $P^I \subseteq \Delta$ and each role name $r \in N_R$ to a binary relation $r^I \subseteq \Delta \times \Delta$. The extension of $\cdot^\mathcal{I}$ to arbitrary concept descriptions is inductively defined, as shown in the third

Construct name	Syntax	Semantics
top-concept	$\top$	Δ
conjunction	$C \sqcap D$	$C^I \cap D^I$
existential restrictions	$\exists r.C$	$\{x \in \Delta \mid \exists y : (x,y) \in r^I \wedge y \in C^I\}$
value restrictions	$\forall r.C$	$\{x \in \Delta \mid \forall y : (x,y) \in r^I \to y \in C^I\}$
primitive negation	$\neg P$	$\Delta \setminus P^I$
bottom-concept	$\bot$	$\emptyset$
number restrictions	$(\geq n\ r)$	$\{x \in \Delta \mid \#\{y \mid (x,y) \in r^{\mathcal{I}}\} \geq n\}$
number restrictions	$(\leq n\ r)$	$\{x \in \Delta \mid \#\{y \mid (x,y) \in r^{\mathcal{I}}\} \leq n\}$

Table 1: Syntax and semantics of concept descriptions. P denotes a concept name from N_C, r a role name from N_R, and n a nonnegative integer.

column of Table 1.

Definition 1 (Subsumption, lcs) *Let C, D, E be concept descriptions of the same DL $\mathcal{L}$.*

1. *D* subsumes *C (for short $C \sqsubseteq D$) iff $C^I \subseteq D^I$ for all interpretations $\mathcal{I}$. C is* equivalent *to D (for short $C \equiv D$) iff $C \sqsubseteq D$ and $D \sqsubseteq C$.*

2. *The concept description E is called* least common subsumer *of C and D iff $C \sqsubseteq E$ and $D \sqsubseteq E$ and for all E' with $C \sqsubseteq E'$ and $D \sqsubseteq E'$, it is $E \sqsubseteq E'$.*

For $\mathcal{ALE}$, subsumption can be characterized by means of *homomorphisms* between *$\mathcal{ALE}$-description trees* [3].

Definition 2 ($\mathcal{ALE}$-description trees) *An $\mathcal{ALE}$-*description tree *is a tree of the form $\mathcal{G} = (V, E, v_0, \ell)$ with root v_0 where*

- *the edges in E are labeled with role names $r \in N_R$ or $\forall r$ for some $r \in N_R$, and*
- *the nodes $v \in V$ are labeled with sets $\ell(v) = \{P_1, \ldots, P_n\}$ where each P_i, $1 \leq i \leq n$, is of one of the following forms: $P_i \in N_C$, $P_i = \neg P$ for some $P \in N_C$, or $P_i = \bot$.*

The empty label corresponds to the top-concept.

Every $\mathcal{ALE}$-concept description can be written (modulo equivalence) as

$$C = Q_1 \sqcap \ldots \sqcap Q_n \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_m.C_m \sqcap \forall s_1.D_1 \sqcap \ldots \sqcap \forall s_k.D_k$$

with $Q_i \in N_C \cup \{\neg P \mid P \in N_C\} \cup \{\bot, \top\}$. The depth $\mathsf{depth}(C)$ of C is defined as the maximal depth of nested quantification in C. Now, the $\mathcal{ALE}$-description tree $\mathcal{G}(C) := (V, E, v_0, \ell)$ corresponding to the $\mathcal{ALE}$-concept description C is inductively defined as follows:

- If $\mathsf{depth}(C) = 0$, then $V := \{v_0\}$, $E := \emptyset$, and $\ell(v_0) := \{Q_1, \ldots, Q_n\} \setminus \{\top\}$;
- otherwise, let $\mathcal{G}_i = (V_i, E_i, v_{0i}, \ell_i)$ be the recursively defined description trees corresponding to C_i, $1 \leq i \leq m$, and $\mathcal{H}_j = (W_j, F_j, w_{0j}, \kappa_j)$ the recursively defined description trees corresponding to D_j, $1 \leq j \leq k$ such that the sets V_i, W_j and $\{v_0\}$ are pairwise disjoint; then
  - $V := \{v_0\} \bigcup_{1 \leq i \leq m} V_i \cup \bigcup_{1 \leq j \leq k} W_j$,
  - $E := \{v_0 r_i v_{0i} \mid 1 \leq i \leq m\} \cup \{v_0 \forall s_j w_{0j} \mid 1 \leq j \leq k\} \cup \bigcup_{1 \leq i \leq m} E_i \cup \bigcup_{1 \leq j \leq k} F_j$, and
  - $\ell(v) := \begin{cases} \{Q_1, \ldots, Q_n\} \setminus \{\top\}, & v = v_0 \\ \ell_i(v), & v \in V_i, 1 \leq i \leq m \\ \kappa_j(v), & v \in W_j, 1 \leq j \leq k. \end{cases}$

The depth of a description tree $\mathcal{G}$ is defined as the length of the longest path in $\mathcal{G}$. Now, each $\mathcal{ALE}$-description tree $\mathcal{G} = (V, E, v_0, \ell)$ is inductively translated into an $\mathcal{ALE}$-concept description $C_\mathcal{G}$ as follows:

- If $\mathsf{depth}(\mathcal{G}) = 0$, then $V = \{v_0\}$ and $E = \emptyset$. Define

$$C_\mathcal{G} := \begin{cases} Q_1 \sqcap \ldots \sqcap Q_n, & \ell(v_0) = \{Q_1, \ldots, Q_n\} \\ \top, & \ell(v_0) = \emptyset; \end{cases}$$

- otherwise, let $\ell(v_0) = \{Q_1, \ldots, Q_n\}$, $n \geq 0$, and let $\{v_1, \ldots, v_m\}$ be the set of all successors of v_0 with $v_0 r_i v_i \in E$ and $\{w_1, \ldots, w_k\}$ the set of all successors of v_0 with $v_0 \forall s_j w_j \in E$. Further, let C_i (D_j) denote the recursively defined concept descriptions obtained from the subtrees with root v_i, $1 \leq i \leq m$ (w_j, $1 \leq j \leq k$). Define

$$C_\mathcal{G} := Q_1 \sqcap \ldots \sqcap Q_n \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_m.C_m \sqcap \forall s_1.D_1 \sqcap \ldots \sqcap \forall s_k.D_k.$$

Definition 3 (Homomorphisms between $\mathcal{ALE}$-description trees)
A homomorphism *from an $\mathcal{ALE}$-description tree $\mathcal{H} = (V_H, E_H, w_0, \ell_H)$ to an $\mathcal{ALE}$-description tree $\mathcal{G} = (V_G, E_G, v_0, \ell_G)$ is a mapping $\varphi : V_H \longrightarrow V_G$ such that*

1. *$\varphi(w_0) = v_0$,*
2. *for all $v \in V_H$ we have $\ell_H(v) \subseteq \ell_G(\varphi(v))$ or $\ell_G(\varphi(v)) = \{\bot\}$,*
3. *for all $vrw \in E_H$, either $\varphi(v) r \varphi(w) \in E_G$, or $\varphi(v) = \varphi(w)$ and $\ell_G(\varphi(v)) = \{\bot\}$, and*
4. *for all $v \forall r w \in E_H$, either $\varphi(v) \forall r \varphi(w) \in E_G$, or $\varphi(v) = \varphi(w)$ and $\ell_G(\varphi(v)) = \{\bot\}$.*

In order to obtain a sound and complete characterization of subsumption, the concept descriptions must be transformed into certain normal forms before translating them into description trees [3]. The *$\mathcal{ALE}$-normal form* of an $\mathcal{ALE}$-concept description C is obtained from C by exhaustively applying the following normalization rules (modulo commutativity and associativity of conjunction):

$$\begin{array}{rrcl}
1. & \forall r.\top & \longrightarrow & \top \\
2. & E \sqcap \top & \longrightarrow & E \\
3. & \forall r.E \sqcap \forall r.F & \longrightarrow & \forall r.(E \sqcap F) \\
4. & \forall r.E \sqcap \exists r.F & \longrightarrow & \forall r.E \sqcap \exists r.(E \sqcap F) \\
5. & P \sqcap \neg P & \longrightarrow & \bot, \text{ for each } P \in N_C \\
6. & \exists r.\bot & \longrightarrow & \bot \\
7. & E \sqcap \bot & \longrightarrow & \bot.
\end{array}$$

The *$\top$-normal form* of C is obtained from C by exhaustively applying th normalization rules 1. and 2. C is said to be in $\mathcal{ALE}$-normal form ($\top$-normal form) if none of the rules (neither rule 1. nor 2.) is applicable to C. Note that the $\mathcal{ALE}$-normal form C' obtained from an $\mathcal{ALE}$-concept description C can be of exponential size w.r.t. the size of C (see [3] for an example). It is easy to see that the $\mathcal{ALE}$-/$\top$-normal form of C is equivalent to C.

Now, subsumption $C \sqsubseteq D$ can be characterized as follows:

Theorem 4 *[3] Let C, D be $\mathcal{ALE}$-concept descriptions, C' the $\mathcal{ALE}$-normal form of C, D' the $\top$-normal form of D, and $\mathcal{G}(C'), \mathcal{G}(D')$ the corresponding $\mathcal{ALE}$-description trees. Then $C \sqsubseteq D$ iff there exists a homomorphism from $\mathcal{G}(D')$ to $\mathcal{G}(C')$.*

Example 5 *Consider the $\mathcal{ALE}$ concept descriptions*

$$\begin{array}{rcl}
C & := & \forall r.\exists r.(P \sqcap \neg P) \sqcap \exists s.(P \sqcap \exists r.Q), \\
D & := & \forall r.(\exists r.P \sqcap \exists r.\neg P) \sqcap \exists s.(\forall r.\top \sqcap \exists r.Q).
\end{array}$$

The $\top$-normal form of D is given by $D' = \forall r.(\exists r.P \sqcap \exists r.\neg P) \sqcap \exists s. \sqcap \exists r.Q$, and the $\mathcal{ALE}$-normal form of C is $C' := \forall r.\bot \sqcap \exists s.(P \sqcap \exists r.Q)$. The corresponding $\mathcal{ALE}$-description trees $\mathcal{G}(C')$ and $\mathcal{G}(D')$ are depicted in Figure 1.

If we define the mapping φ such that it maps w_0 onto v_0; w_1, w_2, and w_3 onto v_1; w_4 onto v_2; and w_5 onto v_3, then it is easy to see that φ yields a homomorphism from $\mathcal{G}(D')$ into $\mathcal{G}(C')$. Thus, Theorem 4 implies $C \sqsubseteq D$.

The inference problem of computing the lcs of $n \geq 2$ concept descriptions has thoroughly been investigated for the DLs $\mathcal{ALE}$ [3], $\mathcal{ALN}$ [5, 2], and Classic [13, 11]. As shown in [3], the lcs of $n \geq 2$ $\mathcal{ALE}$-concept descriptions always exists, and it can be computed in exponential time. The lcs of $n \geq 2$ $\mathcal{ALN}$-concept descriptions also always exists, and it can be computed in polynomial time [5]. Even in the presence of cyclic $\mathcal{ALN}$-concept descriptions, the lcs always exists, and it can be computed in double-exponential time [2]. Things become less rosy, however, if we consider the most specific concept of ABox individuals.

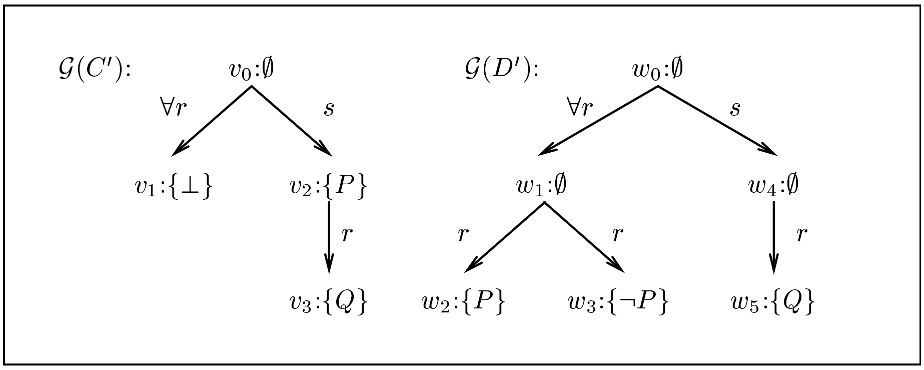

Figure 1: An example for the characterization of subsumption in $\mathcal{ALE}$ by homomorphisms and description trees.

Definition 6 (ABox) *An* ABox $\mathcal{A}$ *is a finite set of assertions of the form* $(a, b) : r$ *(*role assertions*) or* $a : C$ *(*concept assertions*), where* a, b *are* individuals *from a set* N_I*,* r *is a role name, and* C *is a concept description. An ABox is called* $\mathcal{L}$*-ABox if all concept descriptions occurring in* $\mathcal{A}$ *are* $\mathcal{L}$*-concept descriptions.*

In the presence of an ABox, an interpretation $\mathcal{I}$ additionally assigns an element $a^{\mathcal{I}} \in \Delta$ to each individual a occurring in $\mathcal{A}$ such that $A \neq B$ implies $A^{\mathcal{I}} \neq b^{\mathcal{I}}$ (unique name assumption). It is a *model* of $\mathcal{A}$ iff it satisfies $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ for all role assertions $(a, b) : r \in \mathcal{A}$, and $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all concept assertions $a : C \in \mathcal{A}$.

Definition 7 (Instance, msc) *Let* $\mathcal{A}$ *be an* $\mathcal{L}$*-ABox,* a *an individual in* $\mathcal{A}$*, and* C *an* $\mathcal{L}$*-concept description.*

1. a *is an* instance *of* C *w.r.t.* $\mathcal{A}$ *(*$a \in_{\mathcal{A}} C$*) iff* $a^{\mathcal{I}} \in C^{\mathcal{I}}$ *for all models* $\mathcal{I}$ *of* $\mathcal{A}$*.*

2. C *is the* most specific concept *for* a *w.r.t.* $\mathcal{A}$ *iff* $a \in_{\mathcal{A}} C$ *and for all* C' *with* $a \in_{\mathcal{A}} C'$*, it is* $C \sqsubseteq C'$*.*

Depending on the expressive power of the underlying DL $\mathcal{L}$, the msc of an individual a w.r.t. an $\mathcal{L}$-ABox $\mathcal{A}$ need not exist in general. Due to cyclic dependencies between individuals, i.e., cycles build by role assertions in the ABox, it might be the case that there exist infinite many $\mathcal{L}$-concept descriptions a is an instance of, but none most specific concept description with this property. The following example illustrates this situation for $\mathcal{ALN}$ and $\mathcal{ALE}$.

Example 8 *First, consider the* $\mathcal{ALN}$*-ABox* $\mathcal{A} = \{a : P, a : (\leq 1\ r), (a, a) : r\}$*. In [2] it is shown that there does not exist the msc of* a *w.r.t.* $\mathcal{A}$*: It is easy to see that, for each* $n \geq 0$*,* a *is an instance of the* $\mathcal{ALN}$*-concept description*

$$C_n \;:=\; \underbrace{\forall r. \cdots \forall r}_{n\ times}.(P \sqcap (\leq 1\ r) \sqcap (\geq 1\ r)).$$

Intuitively, the msc is thus given as the infinite conjunction $\bigsqcap_{n\geq 0} C_n$. *Obviously, this conjunction cannot be represented by an* $\mathcal{ALN}$*-concept description.*

For $\mathcal{ALE}$ *and its sublanguages* $\mathcal{EL}$ *and* $\mathcal{EL}_\neg$*, we encounter the same problem for the even smaller ABox* $\mathcal{A}' = \{a : P, (a,a) : r\}$*: It is easy to see that, for each* $n \geq 0$*,* a *is an instance of the* $\mathcal{ALE}$*/-*$\mathcal{EL}$*-/*$\mathcal{EL}_\neg$*-concept description*

$$C_n := \underbrace{\exists r. \cdots \exists r}_{n\ times}.P.$$

Assume that there exists an $\mathcal{ALE}$*-/*$\mathcal{EL}$*-/*$\mathcal{EL}_\neg$*-concept description* $C \equiv \mathsf{msc}_{\mathcal{A}'}(a)$*. Let* $\mathsf{depth}(C) = k$*,* C' *the* $\mathcal{ALE}$*-normal form of* C*, and* $\mathcal{G}(C')$ *the corresponding* $\mathcal{ALE}$*-description tree. Obviously, there does not exist a homomorphism from* $\mathcal{G}(C_{k+1})$ *into* $\mathcal{G}(C')$*. Since* C_{k+1} *is in* $\top$*-normal form, Theorem 4 implies* $C \not\sqsubseteq C_{k+1}$ *in contradiction to* $a \in_{\mathcal{A}'} C_{k+1}$ *and* $C \equiv \mathsf{msc}_{\mathcal{A}'}(a)$*.*

For cyclic $\mathcal{ALN}$-ABoxes, the msc can be characterized by *cyclic* $\mathcal{ALN}$-concept descriptions. Such concepts are defined by means of a cyclic TBox, and they are interpreted using the greatest fixed-point semantics [2, 12]. In the above example, the concept C defined by the cyclic $\mathcal{ALN}$-TBox

$$\mathcal{T} = \{C \doteq P \sqcap (\leq 1\, r) \sqcap (\geq 1\, r) \sqcap \forall r.C\}$$

yields the msc of a [2].

As already mentioned in the introduction, cyclic concept descriptions are not yet well-investigated for DLs with existential restrictions. Thus, in this work, we concentrate on approximations of the msc in $\mathcal{ALE}$ (and the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$). The approximation of the msc by concept descriptions with limited depth as introduced in [7] is formally defined as follows:

Definition 9 *Let* $\mathcal{A}$ *be an* $\mathcal{L}$*-ABox,* a *an individual in* $\mathcal{A}$*,* C *an* $\mathcal{L}$*-concept description, and* $k \in \mathbb{N}$ *a nonnegative integer.* C *is called* k-approximation *of* a *w.r.t.* $\mathcal{A}$ *(*$C = \mathsf{msc}_{k,\mathcal{A}}(a)$*) iff*

1. $a \in_{\mathcal{A}} C$,
2. $\mathsf{depth}(C) \leq k$, *and*
3. *for all* C' *with* $a \in_{\mathcal{A}} C'$ *and* $\mathsf{depth}(C') \leq k$*, it is* $C \sqsubseteq C'$*.*

For the ABox $\mathcal{A}'$ from Example 8, the $\mathcal{ALE}$-concept description C_k yields the k-approximation of a w.r.t. $\mathcal{A}$ for each $k \geq 0$.

In the following sections, we will show that, for the DLs $\mathcal{EL}$, $\mathcal{EL}_\neg$, and $\mathcal{ALE}$, the k-approximation of an individual a w.r.t. $\mathcal{A}$ always exists. For $\mathcal{ALE}$, however, we only have a very inefficient algorithm, since the characterization of instance relationships underlying the more efficient algorithms introduced for $\mathcal{EL}$ and $\mathcal{EL}_\neg$ could not be adapted to $\mathcal{ALE}$.

3 Most specific concepts in $\mathcal{EL}$

First, we introduce the characterization of instance in $\mathcal{EL}$ that will be used to prove soundness and completeness of the approximation algorithm presented in Section 3.2.

3.1 Characterizing instance in $\mathcal{EL}$

The characterization of instance can be seen as an extension of the characterization of subsumption given in Theorem 4 (see also[3]). Roughly speaking, the idea is to translate the ABox $\mathcal{A}$ into a so-called $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$, and then to characterize $a \in_{\mathcal{A}} C$ by the existence of a homomorphism φ from the $\mathcal{EL}$-description tree $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ such that the root of $\mathcal{G}(C)$ is mapped onto a.

Definition 10 ($\mathcal{EL}$-description tree/$\mathcal{EL}$-description graph)
An $\mathcal{EL}$-description tree *(*$\mathcal{EL}$-description graph*) is a tree (graph) of the form* $\mathcal{G} = (V, E, v_0, \ell)$ *with root* v_0 *(*$\mathcal{G}(V, E, \ell)$*) where*

- *the edges in* E *are labeled with role names* $r \in N_R$*, and*
- *the nodes* $v \in V$ *are labeled with subsets of* N_C*, i.e.,* $\ell(v) \subseteq N_C$ *for all* $v \in V$*.*

The empty label corresponds to the top-concept.

Note that $\mathcal{EL}$-concept descriptions trivially satisfy the conditions on the $\top$- and $\mathcal{ALE}$-normal form. Thus, Theorem 4 yields

Corollary 11 *Let* C, D *be* $\mathcal{EL}$*-concept descriptions and* $\mathcal{G}(C), \mathcal{G}(D)$ *the corresponding* $\mathcal{EL}$*-description trees. Then,* $C \sqsubseteq D$ *iff there exists a homomorphism* φ *from* $\mathcal{G}(D)$ *into* $\mathcal{G}(C)$*.*

The graphical representation of an $\mathcal{EL}$-ABox $\mathcal{A}$ yields the starting point for the definition of the $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ corresponding to $\mathcal{A}$, i.e., the individuals in $\mathcal{A}$ yield a subset of the nodes in $\mathcal{G}(\mathcal{A})$ and the role assertions in $\mathcal{A}$ yield some of the edges in $\mathcal{G}(\mathcal{A})$. Concept assertions $a : C \in \mathcal{A}$ are translated as follows: the concept names occurring on the top-level of C yield the label of a, and for each existential restriction $\exists r.C'$ on the top-level of C, the $\mathcal{EL}$-description tree $\mathcal{G}(C') = (V_{C'}, E_{C'}, v_{0C'}, \ell_{C'})$ is added to the graph together with the edge $arv_{0C'}$. For example, the $\mathcal{EL}$-ABox

$$\begin{aligned}\mathcal{A} \;=\; & \{a : P \sqcap \exists s.(Q \sqcap \exists r.P \sqcap \exists s.\top), b : P \sqcap Q, c : \exists r.P, \\ & (a, b) : r, (a, c) : r, (b, c) : s\}\end{aligned}$$

yields the $\mathcal{EL}$-description graph depicted in Figure 2. Formally, $\mathcal{G}(\mathcal{A})$ is defined as follows:

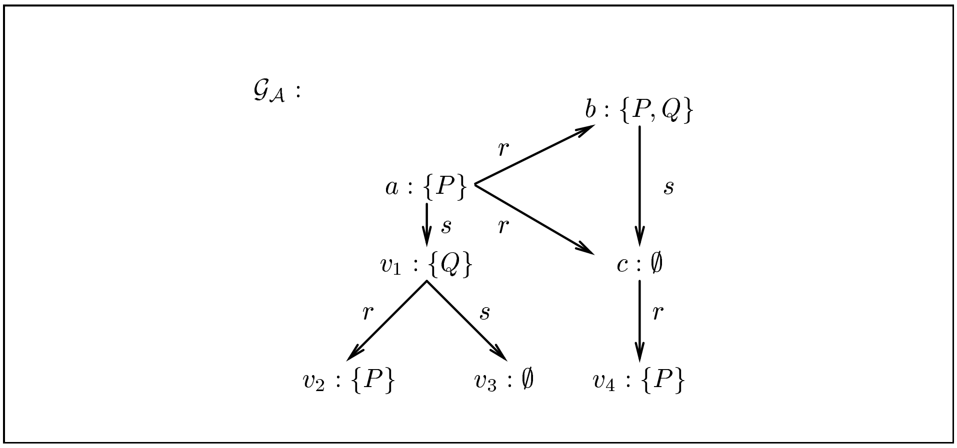

Figure 2: The $\mathcal{EL}$-description graph of $\mathcal{A}$.

Definition 12 ($\mathcal{EL}$-description graph of an $\mathcal{EL}$-ABox)
Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox and $\mathsf{Ind}(\mathcal{A})$ the set of individuals occurring in $\mathcal{A}$. For each $a \in \mathsf{Ind}(\mathcal{A})$, let $C_a := \bigsqcap_{a:D\in\mathcal{A}} D$, if there exists a concept assertion $a : D \in \mathcal{A}$; otherwise, $C_a := \top$. Finally, $\mathcal{G}(C_a) = (V_a, E_a, a, \ell_a)$ denotes the $\mathcal{EL}$-description trees corresponding to C_a, $a \in \mathsf{Ind}(\mathcal{A})$, where w.l.o.g. the sets V_a are assumed to be pairwise disjoint.

The $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ of $\mathcal{A}$ *is defined by $\mathcal{G}(\mathcal{A}) := (V, E, \ell)$ with*

- $V := \bigcup_{a\in\mathsf{Ind}(\mathcal{A})} V_a$,
- $E := \{arb \mid (a, b) : r \in \mathcal{A}\} \cup \bigcup_{a\in\mathsf{Ind}(\mathcal{A})} E_a$, *and*
- $\ell(v) := \ell_a(v)$ *for all* $v \in V_a$.

It remains to adapt the notion of a homomorphism to $\mathcal{EL}$-description graphs and trees: A mapping $\varphi : V_C \longrightarrow V$ is a *homomorphism* from the $\mathcal{EL}$-description tree $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ into the $\mathcal{EL}$-description graph $\mathcal{G} = (V, E, \ell)$ if

1. $\ell_C(v) \subseteq \ell(\varphi(v))$ for all $v \in V_C$, and
2. $\varphi(v)r\varphi(w) \in E$ for all $vrw \in E_C$.

The first condition of Definition 3 is now directly integrated into the characterization of instance relationships in $\mathcal{EL}$.

Theorem 13 *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$ an individual in $\mathcal{A}$, and C an $\mathcal{EL}$-concept description. Further, let $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$ denote the $\mathcal{EL}$-description graph of $\mathcal{A}$ and $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ the $\mathcal{EL}$-description tree corresponding to C. Then, $a \in_\mathcal{A} C$ iff there exists a homomorphism $\varphi : V_C \longrightarrow V$ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$.*

Proof of the if-direction: As already mentioned in the introduction, this characterization of instance is also sound for $\mathcal{ALE}$ (using appropriate definitions of $\mathcal{ALE}$-description graphs and homomorphisms from $\mathcal{ALE}$-description trees into $\mathcal{ALE}$-description graphs). This soundness result will be formalized in Lemma 30 in Section 5. The if-direction of Theorem 13 is an immediate consequence of this lemma.
Proof of the only-if-direction: For this poof, we need the canonical interpretation induced by an $\mathcal{EL}$-description graph.

Definition 14 (Canonical interpretation) *Let $\mathcal{G} = (V, E, \ell)$ be an $\mathcal{EL}$-description graph. The* canonical interpretation $\mathcal{I}(\mathcal{G})$ *is defined by* $\mathcal{I}(\mathcal{G}) := (\Delta_{\mathcal{I}(\mathcal{G})}, \cdot^{\mathcal{I}(\mathcal{G})})$ *with*

- $\Delta_{\mathcal{I}(\mathcal{G})} := V$,
- $P^{\mathcal{I}(\mathcal{G})} := \{v \in V \mid P \in \ell(v)\}$ *for all* $P \in N_C$, *and*
- $r^{\mathcal{I}(\mathcal{G})} := \{(v, w) \in V \times V \mid (vrw \in E)\}$ *for all* $r \in N_R$.

The following lemma formalizes the important property used in the proof of completeness, namely that, for an $\mathcal{EL}$-ABox, the canonical interpretation $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ is a model of $\mathcal{A}$.

Lemma 15 *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox and $\mathcal{G}(\mathcal{A})$ the corresponding $\mathcal{EL}$-description graph. The canonical interpretation $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ is a model of $\mathcal{A}$.*

Proof: We have to show that $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ satisfies each assertion in $\mathcal{A}$.

By construction, $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ satisfies each role assertion in $\mathcal{A}$.

Let $a : D \in \mathcal{A}$ where $D = P_1 \sqcap \ldots \sqcap P_n \sqcap \exists r_1.D_1 \sqcap \ldots \sqcap \exists r_m.D_m$, $P_i \in N_C$. We show that

1. $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in P_i^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq i \leq n$, and
2. $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in (\exists r_j.D_j)^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq j \leq m$.

Ad (1): By definition of $\mathcal{G}(\mathcal{A})$, it is $P_i \in \ell(a)$, and by definition of $\mathcal{I}(\mathcal{G}(\mathcal{A}))$, we get $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in P_i^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq i \leq n$.

Ad (2): Let C_a be defined as in Definition 12. By definition of $\mathcal{G}(\mathcal{A})$, for each $1 \leq j \leq m$, there exists a node $v_j \in V$ such that $ar_jv_j \in E$ and $D_j \equiv C_{\mathcal{G}(C_a)(v_j)}$, where $\mathcal{G}(C_a)(v_j)$ denotes the subtree with root v_j of $\mathcal{G}(C_a)$. By induction on the depth of D_j, it is easy to see that $v_j^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in (C_{\mathcal{G}(C_a)(v_j)})^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$. In addition, it is $(a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}, v_j^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}) \in r_j^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$, and thus $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in (\exists r_j.D_j)^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$.

Now, (1) and (2) imply $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in D^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$. □

In order to complete the proof of the only-if-direction, we will show the following

Claim: Let $v \in V$. If $v \in D^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$, then there exists a homomorphism φ from $\mathcal{G}(D) = (V_D, E_D, w_0, \ell_D)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = v$.

Since $a \in_{\mathcal{A}} C$ and $\mathcal{I}(\mathcal{G}(\mathcal{A})) \models \mathcal{A}$, the claim implies that there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$.

We prove the claim by induction on $\mathsf{depth}(D)$.

$\mathsf{depth}(D) = 0$, i.e., $D = P_1 \sqcap \ldots \sqcap P_n$. Then $\mathcal{G}(D) = (\{w_0\}, \emptyset, w_0, \ell_D)$ with $\ell_D(w_0) = \{P_1, \ldots, P_n\}$. Define φ by $\varphi(w_0) := v$. Since $v \in D^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$, the definition of $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ implies $\ell_D(w_0) \subseteq \ell(v)$, i.e., φ is a homomorphism from $\mathcal{G}(D)$ into $\mathcal{G}(\mathcal{A})$.

$\mathsf{depth}(D) > 0$, i.e., $D = P_1 \sqcap \ldots \sqcap P_n \sqcap \exists r_1.D_1 \sqcap \ldots \sqcap \exists r_m.D_m$. As for $\mathsf{depth}(D) =)$ we get $\{P_1, \ldots, P_n\} \subseteq \ell(v)$. Now, $v \in D^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ implies that, for each $1 \leq j \leq m$, there exists a node $v_i \in V$ such that $(v, v_i) \in r^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ and $v_i \in D_m^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$. Let w_i denote the r_i-successor of w_0 in $\mathcal{G}(D)$ with $D_i = \mathcal{G}(D)(w_i)$, $1 \leq i \leq m$. By induction, there exist homomorphisms $\varphi_{w_1}, \ldots, \varphi_{w_m}$ from $\mathcal{G}(D)(w_i)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi_{w_i}(w_i) = v_i$ for all $1 \leq i \leq m$. Define φ by

$$\varphi := \{w_0 \mapsto v\} \cup \bigcup_{w_0 r w \in E_D} \varphi_w.$$

Since for each $w_0 r w \in E_D$ there exists an $j \in \{1, \ldots, m\}$ such that $w = w_j$, φ is well-defined, and by construction, it is a homomorphism from $\mathcal{G}(D)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = v$. □

Whether there exists a homomorphism from a tree into a graph can be decided in polynomial time [8]. Since $\mathcal{G}(C)$ and $\mathcal{G}(\mathcal{A})$ can be computed in polynomial time, we get

Proposition 16 *The instance problem for $\mathcal{EL}$ can be decided in polynomial time.*

3.2 Computing k-approximations in $\mathcal{EL}$

In this section, we will show that, for an $\mathcal{EL}$ABox $\mathcal{A}$ and an individual $a \in \mathsf{Ind}(\mathcal{A})$, the k-approximation of a w.r.t. $\mathcal{A}$ always exists and can be effectively computed. The algorithm computing $\mathsf{msc}_{k,\mathcal{A}}(a)$ introduced below works as follows: First, the description graph $\mathcal{G}(\mathcal{A})$ is unraveled into a tree $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ with root a, a finite branching factor, but possibly infinite long paths. Truncating all paths of length $\leq k$ then yields an $\mathcal{EL}$-description tree $\mathcal{T}_k(a, \mathcal{A})$ of depth $\leq k$. Using the characterization of subsumption introduced above, it is easy to show that the $\mathcal{EL}$-concept description $C_{\mathcal{T}_k(\mathcal{G}(\mathcal{A}))}$ is equivalent to $\mathsf{msc}_{k,\mathcal{A}}(a)$. In case that $\mathcal{A}$ is acyclic, $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is an $\mathcal{EL}$-description tree and $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ yields the msc of a w.r.t. $\mathcal{A}$.

For the definition of the trees $\mathcal{T}(a, \mathcal{A})$ and $\mathcal{T}_k(a, \mathcal{A})$, we need the following notions: For an $\mathcal{EL}$-description graph $\mathcal{G} = (V, E, \ell)$, $p = v_0 r_1 v_1 r_2 \ldots r_n v_n$ is a *path from v_0 to v_n* of *length* $|p| = n$, if $v_{i-1} r_i v_i \in E$ for all $1 \leq i \leq n$. The path p will also be denoted as $r_1 \cdots r_n$*-path from v_0 to v_n*, and the node v_n as $r_1 \cdots r_n$*-successor of* v_0, whereby each node is assumed to be an ε-successor of itself.

The path p contains a *cycle*, if $v_i = v_j$ for two indices i, j with $0 \leq i < j \leq n$. A node v is *reachable* from v_0, if there exists a path from v_0 to v.

Definition 17 (Tree of a w.r.t. $\mathcal{G}$ (and k)) *Let $\mathcal{G} = (V, E, \ell)$ and $a \in V$. The* tree $\mathcal{T}(a, \mathcal{G})$ of a w.r.t. $\mathcal{A}$ *is defined by $\mathcal{T}(a, \mathcal{G}) := (V^t, E^t, a, \ell^t)$ with*

- $V^t := \{ar_1v_1r_2 \dots r_nv_n \mid ar_1v_1r_2 \dots r_nv_n$ *is a path from a to v_n in* $\mathcal{G}\}$,
- $E^t := \{prq \mid p, q \in V^t$ *and* $p = ar_1v_1r_2 \dots r_nv_n$ *and* $q = ar_1v_1r_2 \dots r_nv_nrw\}$,
- $\ell^t(p) := \ell(v)$ *if* $p = ar_1v_1r_2 \dots r_nv$.

For a nonnegative integer $k \in \mathbb{N}$ the tree $\mathcal{T}_k(a, \mathcal{G})$ of a w.r.t. $\mathcal{G}$ and k *is defined by $\mathcal{T}_k(a, \mathcal{G}) := (V^t_k, E^t_k, a, \ell^t_k)$ with*

- $V^t_k := \{p \in V^t \mid |p| \leq k\}$,
- $E^t_k := E^t \cap (V^t_k \times N_R \times V^t_k)$, *and*
- $\ell^t_k(p) := \ell^t(p)$ *if* $p \in V^t_k$.

Since by definition $\mathcal{EL}$-description graphs are finite, $\mathcal{T}(a, \mathcal{G})$ has a finite branching factor, i.e., $\mathcal{T}(a, \mathcal{G})$ is infinite if and only if $\mathcal{T}(a, \mathcal{G})$ contains a path of infinite length. This is the case if and only if there exists a path p in $\mathcal{G}$ such that p contains a cycle. By definition, $\mathcal{T}_k(a, \mathcal{G})$ is an $\mathcal{EL}$-description tree of depth $\leq k$. If $\mathcal{G}$ is acyclic, i.e., $\mathcal{G}$ does not exist a path p in $\mathcal{G}$ containing a cycle, $\mathcal{T}(a, \mathcal{G})$ is finite and thus, an $\mathcal{EL}$-description tree. Now, the following characterization of the (k-approximation of the) msc is based on these trees, whereby the $\mathcal{EL}$-description graph is obtained from an $\mathcal{EL}$-ABox $\mathcal{A}$, and the root is assumed to be an individual from $\mathcal{A}$.

Theorem 18 *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$. Then, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is the k-approximation of a w.r.t. $\mathcal{A}$. If, starting from a, no cyclic path in $\mathcal{A}$ can be reached (i.e., $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is finite), then $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ is the msc of a w.r.t. $\mathcal{A}$; otherwise no msc exists.*

Proof: Let $\mathcal{G}(\mathcal{A}) = (V_\mathcal{A}, E_\mathcal{A}, \ell_\mathcal{A})$ and $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A})) = (V_k, E_k, a, \ell_k)$.

We first show $a \in_\mathcal{A} C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$. Let φ be the mapping obtained from mapping each path $p \in V_k$ onto the last node v occurring in p. It is easy to see that φ yields a homomorphism from $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(a) = a$. Since, for $\mathcal{EL}$-description trees $\mathcal{G}$, it is $\mathcal{G} = \mathcal{G}(C_\mathcal{G})$ up to renaming nodes, Theorem 13 yields $a \in_\mathcal{A} C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$.

It remains to show that for all $\mathcal{EL}$-concept descriptions C with $a \in_\mathcal{A} C$ and $\mathsf{depth}(C) \leq k$ also $C_{\mathcal{T}_k(a,\mathcal{A}))} \sqsubseteq C$. Let C be such an $\mathcal{EL}$-concept description. By Theorem 13 we get that there exists a homomorphism φ from $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$. Using φ, we define a homomorphism ψ from $\mathcal{G}(C)$ into $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$. Then Theorem 4 implies $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))} \sqsubseteq C$.

For $v \in V_C$, let $v_0r_1v_1r_2 \cdots r_{n-1}v_{n-1}r_nv$ be the unique (!) path from v_0 to v in $\mathcal{G}(C)$. Define $p(v) := \varphi(v_0)r_1\varphi(v_1)r_2 \cdots r_{n-1}\varphi(v_{n-1})r_n\varphi(v)$. Since φ is a homomorphism from $\mathcal{G}(C)$ into $G(\mathcal{A})$, $p(v)$ is well-defined and yields a path of length $n \leq k$ from $\varphi(v_0)$ to $\varphi(v)$ in $\mathcal{G}(\mathcal{A})$. It is easy to see that the mapping $\psi : V_C \longrightarrow V_k$ defined by $\psi(v) := p(v)$ yields a homomorphism from $\mathcal{G}(C)$ into $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$.

Now, assume that, starting from a, a cycle can be reached in $\mathcal{A}$, i.e., $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is infinite. Then, we have a decreasing chain $C_0 \sqsupset C_1 \sqsupset \cdots$ of k-approximations C_k ($\equiv C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$) with increasing depth k, $k \geq 0$. From Theorem 11, we conclude that there does not exist an $\mathcal{EL}$-concept description subsumed by all of these k-approximations (since such a concept description only has a fixed and finite depth). Thus, a cannot have an msc.

Conversely, if $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is finite, say with depth k, from the observation that all k'-approximations, for $k' \geq k$, are equivalent, it immediately follows that $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ is the msc of a. □

Obviously, there exists a deterministic algorithm computing the k-approximation (i.e., $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$) in time $\mathcal{O}(|\mathcal{A}|^k)$. The *size* $|\mathcal{A}|$ of $\mathcal{A}$ is defined by

$$|\mathcal{A}| := |\mathsf{Ind}(\mathcal{A})| + |\{(a,b) : r \mid (a,b) : r \in \mathcal{A}\}| + \sum_{a:C \in \mathcal{A}} |C|,$$

where the size $|C|$ of C is defined as the sum of the number of occurrences of concept names, role names, and constructors in C. Similarly, one obtains an exponential complexity upper bound for computing the msc (if it exists).

Corollary 19 *For an $\mathcal{EL}$-ABox $\mathcal{A}$, an individual $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$, the k-approximation of a w.r.t. $\mathcal{A}$ always exists and can be computed in time $\mathcal{O}(|\mathcal{A}|^k)$.*

The msc of a exists iff starting from a no cycle can be reached in $\mathcal{A}$. The existence of the msc can be decided in polynomial time, and if the msc exists, it can be computed in time exponential in the size of $\mathcal{A}$.

In the remainder of this section, we prove that the exponential upper bounds are tight. To this end, we show examples demonstrating that k-approximations and the msc may grow exponentially.

Example 20 *Let $\mathcal{A} = \{(a,a) : r, (a,a) : s\}$. The $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ as well as the $\mathcal{EL}$-description trees $\mathcal{T}_1(a, \mathcal{G}(\mathcal{A}))$ and $\mathcal{T}_2(a, \mathcal{G}(\mathcal{A}))$ are depicted in Figure 3. It is easy to see that, for $k \geq 1$, $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ yields a full binary tree of depth k where*

- *each node is labeled with the empty set, and*
- *each node except the leaves has one r- and one s-successor.*

By Theorem 18, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is the k-approximation of a w.r.t. $\mathcal{A}$. The size of $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is $|\mathcal{A}|^k$. Moreover, it is not hard to see that there does not exist an $\mathcal{EL}$-concept description C which is equivalent to but smaller than $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$.

The following example illustrates that, if it exists, also the msc can be of exponential size.

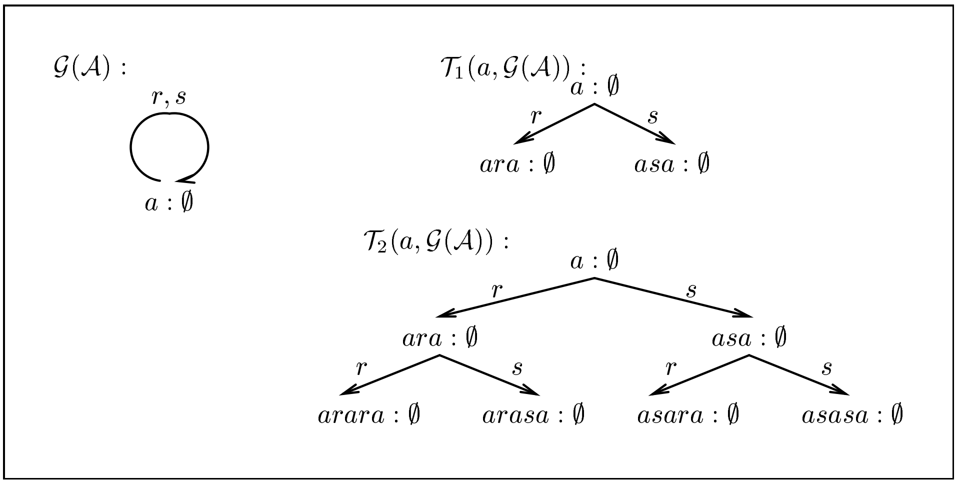

Figure 3: The $\mathcal{EL}$-description graph and the $\mathcal{EL}$-description trees from Example 20.

Example 21 *For $n \geq 1$, define $\mathcal{A}_n := \{(a_i, a_{i+1}) : r, (a_i, a_{i+1}) : s \mid 1 \leq i < n\}$. Obviously, $\mathcal{A}_n$ is acyclic, and the size of $\mathcal{A}_n$ is linear in n. By Theorem 18, $C_{\mathcal{T}(a_1,\mathcal{A}_n)}$ is the msc of a_1 w.r.t. $\mathcal{A}_n$. It is easy to see that, for each n, $\mathcal{T}(a_1, \mathcal{A}_n)$ coincides with the tree $\mathcal{T}_n(a, \mathcal{G}(\mathcal{A}))$ obtained in Example 20. As before we obtain that*

- *$C_{\mathcal{T}(a_1,\mathcal{G}(\mathcal{A}))}$ is of size exponential in $|\mathcal{A}_n|$; and*
- *there does not exist an $\mathcal{EL}$-concept description C equivalent to but smaller than $C_{\mathcal{T}(a_1,\mathcal{G}(\mathcal{A}))}$.*

Summarizing, we obtain the following lower bounds.

Proposition 22 *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$.*

- *The size of $\mathsf{msc}_{\mathcal{A},k}(a)$ may grow with $|\mathcal{A}|^k$.*
- *If it exists, the size of $\mathsf{msc}_{\mathcal{A}}(a)$ may grow exponentially in $|\mathcal{A}|$.*

4 Most specific concepts in $\mathcal{EL}_\neg$

Our goal is to obtain a characterization of the (k-approximation of the) msc in $\mathcal{EL}_\neg$ analogously to the one given in Theorem 18 for $\mathcal{EL}$. To achieve this goal, first the notions description graph and description tree are extended from $\mathcal{EL}$ to $\mathcal{EL}_\neg$ by allowing for subsets of $N_C \cup \{\neg P \mid P \in N_C\} \cup \{\bot\}$ as node labels. Just as for $\mathcal{EL}$, there exists a 1–1 correspondence between $\mathcal{EL}_\neg$-concept descriptions and $\mathcal{EL}_\neg$-description trees, and an $\mathcal{EL}_\neg$-ABox $\mathcal{A}$ is translated into an $\mathcal{EL}_\neg$-description graph $\mathcal{G}(\mathcal{A})$ as described for $\mathcal{EL}$-ABoxes. The notion of a homomorphism also

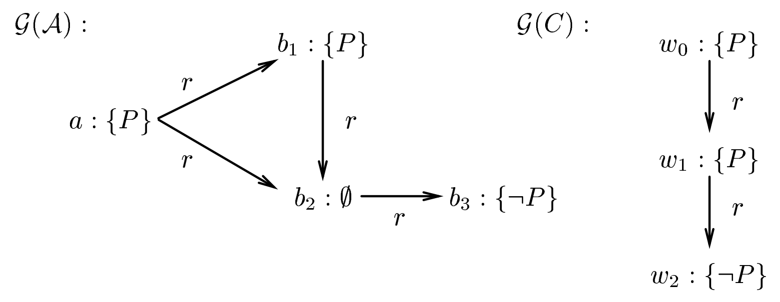

Figure 4: The $\mathcal{EL}_\neg$-description graph and the $\mathcal{EL}_\neg$-description tree from Example 23.

remains unchanged for $\mathcal{EL}_\neg$, and the characterization of subsumption extends to $\mathcal{EL}_\neg$ by just considering inconsistent $\mathcal{EL}_\neg$-concept descriptions as a special case: $C \sqsubseteq D$ iff $C \equiv \bot$ or there exists a homomorphism φ from $\mathcal{G}(D)$ into $\mathcal{G}(C)$.

Second, we have to cope with inconsistent $\mathcal{EL}_\neg$-ABoxes as a special case: for an inconsistent ABox $\mathcal{A}$, $a \in_{\mathcal{A}} C$ is valid for all concept descriptions C, and hence, $\mathsf{msc}_{\mathcal{A}}(a) \equiv \bot$. However, extending Theorem 13 with this special case does not yield a sound and complete characterization of instance relationships for $\mathcal{EL}_\neg$. If this was the case, we would get that the instance problem for $\mathcal{EL}_\neg$ is in P, in contradiction to complexity results shown in [16], which imply that the instance problem for $\mathcal{EL}_\neg$ is coNP-hard.

The following example is an abstract version of an example given in [16]; it illustrates incompleteness of a naïve extension of Theorem 13 from $\mathcal{EL}$ to $\mathcal{EL}_\neg$.

Example 23 *Consider the $\mathcal{EL}_\neg$-concept description $C = P \sqcap \exists r.(P \sqcap \exists r.\neg P)$ and the $\mathcal{EL}_\neg$-ABox $\mathcal{A} = \{a : P, b_1 : P, b_3 : \neg P, (a,b_1) : r, (a,b_2) : r, (b_1,b_2) : r, (b_2,b_3) : r\}$; $\mathcal{G}(\mathcal{A})$ and $\mathcal{G}(C)$ are depicted in Figure 4. Obviously, there does not exist a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = a$, because neither $P \in \ell(b_2)$ nor $\neg P \in \ell(b_2)$. For each model $\mathcal{I}$ of $\mathcal{A}$, however, either $b_2^{\mathcal{I}} \in P^{\mathcal{I}}$ or $b_2^{\mathcal{I}} \in (\neg P)^{\mathcal{I}}$, and in fact, $a^{\mathcal{I}} \in C^{\mathcal{I}}$. Thus, a is an instance of C w.r.t. $\mathcal{A}$ though there does not exist a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = a$.*

In the following section, we give a sound and complete characterization of instance relationships in $\mathcal{EL}_\neg$, which again yields the basis for the characterization of k-approximations given in Section 4.2.

4.1 Characterizing instance in $\mathcal{EL}_\neg$

The reason for the problem illustrated in Example 23 is that, in general, for the individuals in the ABox it is not always fixed whether they are instances of a given concept name or not. Thus, in order to obtain a sound and complete characterization analogous to Theorem 13, instead of $\mathcal{G}(\mathcal{A})$, one has to consider all so-called atomic completions of $\mathcal{G}(\mathcal{A})$.

Definition 24 (Atomic completion) *Let $\mathcal{G} = (V, E, \ell)$ be an $\mathcal{EL}_\neg$-description graph and $N_C^* := \{P \in N_C \mid \text{exists } v \in V \text{ with } P \in \ell(v) \text{ or } \neg P \in \ell(v)\}$. An $\mathcal{EL}_\neg$-description graph $\mathcal{G}^* = (V, E, \ell^*)$ is an* atomic completion *of $\mathcal{G}$ if, for all $v \in V$,*

1. *$\ell(v) \subseteq \ell^*(v)$,*
2. *for all concept names $P \in N_C^*$ either $P \in \ell^*(v)$ or $\neg P \in \ell^*(v)$.*

Note that by definition, all labels of nodes in completions do not contain a conflict, i.e., the nodes are not labeled with a concept name and its negation. In particular, if $\mathcal{G}$ has a conflicting node, then $\mathcal{G}$ does not have a completion. It is easy to see that an $\mathcal{EL}$-ABox $\mathcal{A}$ is inconsistent iff $\mathcal{G}(\mathcal{A})$ contains a conflicting node. For this reason, in the following characterization of the instance relationship, we do not need to distinguish between consistent and inconsistent ABoxes.

Theorem 25 *Let $\mathcal{A}$ be an $\mathcal{EL}_\neg$-ABox, $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$ the corresponding description graph, C an $\mathcal{EL}_\neg$-concept description, $\mathcal{G}(C) = (V_C, E_C, w_0, \ell_C)$ the corresponding description tree, and $a \in \mathsf{Ind}(\mathcal{A})$. Then, $a \in_{\mathcal{A}} C$ iff for each atomic completion $\mathcal{G}(\mathcal{A})^*$ of $\mathcal{G}(\mathcal{A})$, there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$ with $\varphi(w_0) = a$.*

Proof of the if-direction: For the characterization of instance in $\mathcal{EL}$ (see Theorem 13, the proof of the if-direction could be obtained trivially as a special case of the soundness result given for $\mathcal{ALE}$ in Section 5. For $\mathcal{EL}_\neg$, however, things are not that easy: Since for the if-direction we only assume that there exist homomorphisms from $\mathcal{G}(C)$ into *primitive completions* of $\mathcal{G}(\mathcal{A})$, and since a primitive completion of $\mathcal{G}(\mathcal{A})$ in general does not coincide with $\mathcal{G}(\mathcal{A})$, the preconditions of Lemma 30 are not satisfied.

The idea underlying the proof given below is as follows: For a consistent $\mathcal{EL}_\neg$-ABox $\mathcal{A}$ (for inconsistent ABoxes nothing has to be shown) and a model $\mathcal{I} = (\Delta_\mathcal{I}, \cdot^\mathcal{I})$ of $\mathcal{A}$, we first define a mapping $\psi : V \longrightarrow \Delta_\mathcal{I}$ with

1. $\psi(b) = b^\mathcal{I}$ for all $b \in \mathsf{Ind}(\mathcal{A})$,
2. $(\psi(v), \psi(w)) \in r^\mathcal{I}$ for all $vrw \in E$, and
3. $\psi(v) \in Q^\mathcal{I}$ for all $Q \in \ell(v)$ and $v \in V$.

Using this mapping, we then define a primitive completion $\mathcal{G}(\mathcal{A})^*_\psi = (V, E, \ell^*_\psi)$ in such a way that ψ also satisfies condition

4. $\psi(v) \in Q^\mathcal{I}$ for all $Q \in \ell^*_\psi(v)$ and $v \in V$.

By assumption, there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*_\psi$. By induction on the depth of $\mathcal{G}(C)(w)$, we finally show that for all $w \in V_C$

$$\psi(\varphi(w)) \in C_{\mathcal{G}(C)(w)}{}^{\mathcal{I}}. \tag{1}$$

Since $\varphi(w_0) = a$ and $\psi(a) = a^{\mathcal{I}}$ and $C \equiv C_{\mathcal{G}(C)(w_0)}$, this implies $a^{\mathcal{I}} \in C^{\mathcal{I}}$.

The definition of the mapping ψ*:* Let $a \in \mathsf{Ind}(\mathcal{A})$, $C_a := \bigsqcap_{a:D\in\mathcal{A}} D$, and $\mathcal{G}(C_a) = (V_a, E_a, a, \ell_a)$. For $v \in V_a$, we define $\psi(v)$ by induction on the length λ of the unique (!) path from a to v in $\mathcal{G}(C_a)$ such that in addition $\psi(v) \in C_{\mathcal{G}(C_a)(v)}{}^{\mathcal{I}}$.

$\lambda = 0$: Then $v = a$. Define $\psi(a) := a^{\mathcal{I}}$. Since $\mathcal{I} \models \mathcal{A}$, we get $a^{\mathcal{I}} \in C_a^{\mathcal{I}}$. In particular, $a^{\mathcal{I}} \in Q^{\mathcal{I}}$ for all $Q \in \ell(a)$. Thus, conditions (1) and (3) are satisfied for all $a \in \mathsf{Ind}(\mathcal{A})$.

$\lambda > 0$: Then there exists a unique edge of the form $v'rv \in E_a$. By induction, $\psi(v')$ is already defined, and it is $\psi(v') \in C_{\mathcal{G}(C_a)(v')}{}^{\mathcal{I}}$. Since $v'rv \in E_a$, there exists an existential restriction of the form $\exists r.C_{\mathcal{G}(C_a)(v)}$ on the top-level of $C_{\mathcal{G}(C_a)(v')}$. Now, $\psi(v') \in C_{\mathcal{G}(C_a)(v')}{}^{\mathcal{I}}$ implies that there exists an $\alpha \in \Delta_{\mathcal{I}}$ such that $(\psi(v'), \alpha) \in R^{\mathcal{I}}$ and $\alpha \in C_{\mathcal{G}(C_a)(v)}{}^{\mathcal{I}}$. Define $\psi(v) := \alpha$. Since $\alpha \in C_{\mathcal{G}(C_a)(v)}{}^{\mathcal{I}}$ and $\ell(v) = \ell_a(v)$, we get $\alpha \in Q^{\mathcal{I}}$ for all $Q \in \ell(v)$, i.e., condition (3) is satisfied for v.

By construction, ψ satisfies the conditions (1)–(3).

The definition of the primitive completion $\mathcal{G}(\mathcal{A})^*_\psi$*:* Let N_C^* be the set of all concept names occurring in $\mathcal{G}(\mathcal{A})$. Define $\mathcal{G}(\,T)^*_\psi := (V, E, \ell^*_\psi)$ by

$$\ell^*_\psi(v) := \{P \in N_C^* \mid \psi(v) \in P^{\mathcal{I}}\} \cup \{\neg P \mid P \in N_C^* \text{ and } \psi(v) \notin P^{\mathcal{I}}\}.$$

By condition (3) we get $\ell(v) \subseteq \ell^*_\psi(v)$ for all $v \in V$. Thus, $\mathcal{G}(\mathcal{A})^*_\psi$ is a primitive completion of $\mathcal{G}(\mathcal{A})$ that, by definition, satisfies condition (4).

Now, the precondition of the if-direction yields a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*_\psi$ with $\varphi(w_0) = a$. We show Property (1) by induction on the depth of $\mathcal{G}(C)(w)$:

$\mathsf{depth}(\mathcal{G}(C)(w)) = 0$: Then $C_{\mathcal{G}(C)(w)} = \bigsqcap_{Q\in\ell_C(w)} Q$.

Since $\ell_C(w) \subseteq \ell^*_\psi(\varphi(w))$, condition (4) implies $\psi(\varphi(w)) \in Q^{\mathcal{I}}$ for all $Q \in \ell_C(w)$. Hence, $\psi(\varphi(w)) \in C_{\mathcal{G}(C)(w)}{}^{\mathcal{I}}$.

$\mathsf{depth}(\mathcal{G}(C)(w)) > 0$: Then $C_{\mathcal{G}(C)(w)} = \bigsqcap_{Q\in\ell_C(w)} Q \sqcap \bigsqcap_{wrw'\in E_C} \exists r.C_{\mathcal{G}(C)(w')}$.

As before we get $\psi(\varphi(w)) \in Q^{\mathcal{I}}$ for all $Q \in \ell_C(w)$. Let $wrw' \in E_C$. By definition of ψ, and since φ is a homomorphism, we get $(\psi(\varphi(w)), \psi(\varphi(w'))) \in r^{\mathcal{I}}$. By induction, $\psi(\varphi(w')) \in C_{\mathcal{G}(C)(w')}{}^{\mathcal{I}}$, and hence $\psi(\varphi(w)) \in (\exists r.C_{\mathcal{G}(C)(w')})^{\mathcal{I}}$.

Summing up, we get $\psi(\varphi(w)) \in C_{\mathcal{G}(C)(w)}{}^{\mathcal{I}}$.

This completes the proof of the if-direction.

Proof of the only-if-direction: Let $\mathcal{A}$ be a consistent $\mathcal{EL}_{\neg}$-ABox, $N_C^* := \{P \in N_C \mid \exists v \in V : P \in \ell(v) \vee \neg P \in \ell(v)\}$, and $a \in_{\mathcal{A}} C$. Let $\mathcal{G}(C) = (V_C, E_C, w_0, \ell_C)$, $\mathcal{G}(\mathcal{A})^* = (V, E, \ell^*)$ an arbitrary primitive completion of $\mathcal{G}(\mathcal{A})$, and $\mathcal{I}(\mathcal{G}(\mathcal{A})^*)$ the canonical interpretation induced by $\mathcal{G}(\mathcal{A})^*$. The canonical interpretation of an $\mathcal{EL}_{\neg}$-ABox is defined just as for $\mathcal{EL}$. Moreover, it is easy to see that the canonical interpretation of an $\mathcal{EL}_{\neg}$-ABox $\mathcal{A}$ is a model of $\mathcal{A}$. Since

- $r^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} = r^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ for all $r \in N_R$,
- $P^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \subseteq P^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ for all $P \in N_C$, and
- $\neg P \in \ell(v) \implies P \notin \ell^*(v) \implies v \notin P^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$,

$\mathcal{I}(\mathcal{G}(\mathcal{A})^*)$ is also a model of $\mathcal{A}$. We first show that, for all concept names P occurring in C, $P \in N_C^*$. Assume that there exists a concept name P occurring in C with $P \notin N_C^*$. Let $w \in V_C$ be an $r_1 \ldots r_n$-successor of w_0 in $\mathcal{G}(C)$ with $P \in \ell_C(w)$ or $\neg P \in \ell_C(w)$. Assume $P \in \ell_C(w)$. Then, for each model $\mathcal{I}$ of C and each $\alpha \in C^{\mathcal{I}}$, there exists an $(r_1 \ldots r_n)^{\mathcal{I}}$-successor β of α in $\mathcal{I}$ with $\beta \in P^{\mathcal{I}}$. Since $P \notin N_C^*$, however, $P^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)} = \emptyset$, and hence, $C^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)} = \emptyset$ in contradiction to $a \in_{\mathcal{A}} C$. If $\neg P \in \ell_C(w)$, then for each model $\mathcal{I}$ of C and each $\alpha \in C^{\mathcal{I}}$, there exists an $(r_1 \ldots r_n)^{\mathcal{I}}$-successor β of α in $\mathcal{I}$ with $\beta \notin P^{\mathcal{I}}$. Define $\mathcal{J} := (V, \cdot^{\mathcal{J}})$, where $Q^{\mathcal{J}} := Q^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ for all $Q \in N_C^*$ and $P^{\mathcal{J}} := V$. Then $\mathcal{J}$ is a model of $\mathcal{A}$, because the interpretation of concept names and role names occurring in $\mathcal{A}$ remained unchanged w.r.t. $\mathcal{I}(\mathcal{G}(\mathcal{A})^*)$. But obviously, $(\neg P)^{\mathcal{J}} = \emptyset$, and hence $C^{\mathcal{J}} \neq \emptyset$ in contradiction to $a \in_{\mathcal{A}} C$.

Thus, we have shown that, for a consistent $\mathcal{EL}_{\neg}$-ABox $\mathcal{A}$, $a \in_{\mathcal{A}} C$ implies $P \in N_C^*$ for all concept names P occurring in C. Now, the only-if-direction is an easy consequence of the following

Claim: If $v \in C^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ and if all concept names occurring in C also occur in $\mathcal{A}$, then there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$ with $\varphi(w_0) = v$.

Proof by induction on $\mathsf{depth}(C)$:

$\mathsf{depth}(C) = 0$: Then $C = Q_1 \sqcap \ldots \sqcap Q_n$, where $Q_i \in N_C \cup \{\neg P \mid P \in N_C\}$.

Define φ by $\varphi(w_0) := v$. We know $v \in C^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$. If $Q_i \in N_C$, then by definition of $\mathcal{I}(\mathcal{G}(\mathcal{A})^*)$, it is $Q_i \in \ell(v)$. If $Q_i = \neg P$ for some $P \in N_C$, $v \in Q_i^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ implies $P \notin \ell(v)$. The precondition on C yields $P \in N_C^*$. Since $\mathcal{G}(\mathcal{A})^*$ is a primitive completion, we get $\neg P \in \ell(v)$. Summing up, it is $\ell_C(w_0) \subseteq \ell^*(v)$, and hence φ is a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$.

$\mathsf{depth}(C) > 0$: The induction step is shown as for $\mathcal{EL}$.

This completes the proof of the only-if-direction and hence of Theorem 25. □

The problem of deciding whether there exists an atomic completion $\mathcal{G}(\mathcal{A})^*$ such that there exists no homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$ is in coNP. Adding the coNP-hardness result obtained from [16], this shows

Corollary 26 *The instance problem for $\mathcal{EL}_\neg$ is coNP-complete.*

4.2 Computing k-approximations in $\mathcal{EL}_\neg$

Not surprisingly, the algorithm computing the k-approximation/msc in $\mathcal{EL}$ does not yield the desired result for $\mathcal{EL}_\neg$. For instance, in Example 23, we would get $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))} = P \sqcap \exists r.\exists r.(\neg P) \sqcap \exists r.(P \sqcap \exists r.\exists r.(\neg P))$. But as we will see, $\mathsf{msc}_\mathcal{A}(a) \equiv P \sqcap \exists r.(P \sqcap \exists r.\neg P) \sqcap \exists r.(P \sqcap \exists r.\exists r.\neg P)$, i.e., $\mathsf{msc}_\mathcal{A}(a) \sqsubset C_{\mathcal{T}(a,\mathcal{A})}$.

As in the extension of the characterization of instance relationships from $\mathcal{EL}$ to $\mathcal{EL}_\neg$, we have to take into account all atomic completions instead of the single description graph $\mathcal{G}(\mathcal{A})$. Intuitively, one has to compute the least concept description for which there exists a homomorphism into each atomic completion of $\mathcal{G}(\mathcal{A})$. In fact, this can be done by applying the lcs operation on the set of all concept descriptions $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^*)}$ obtained from the atomic completions $\mathcal{G}(\mathcal{A})^*$ of $\mathcal{G}(\mathcal{A})$.

Theorem 27 *Let $\mathcal{A}$ be an $\mathcal{EL}_\neg$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$. If $\mathcal{A}$ is inconsistent, then $\mathsf{msc}_{k,\mathcal{A}}(a) \equiv \mathsf{msc}_\mathcal{A}(a) \equiv \bot$. Otherwise, let $\{\mathcal{G}(\mathcal{A})^1, \ldots, \mathcal{G}(\mathcal{A})^n\}$ be the set of all atomic completions of $\mathcal{G}(\mathcal{A})$.*

Then, $\mathsf{lcs}(C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^1)}, \ldots, C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^n)}) \equiv \mathsf{msc}_{k,\mathcal{A}}(a)$. If, starting from a, no cycle can be reached in $\mathcal{A}$, then $\mathsf{lcs}(C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^1)}, \ldots, C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^n)}) \equiv \mathsf{msc}_\mathcal{A}(a)$; otherwise the msc does not exist.

Proof sketch. Let $\mathcal{A}$ be a consistent $\mathcal{EL}_\neg$-ABox and $\mathcal{G}(\mathcal{A})^1, \ldots, \mathcal{G}(\mathcal{A})^n$ the atomic completions of $\mathcal{G}(\mathcal{A})$. By definition of $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)}$, there exists a homomorphism π_i from $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)}$ into $\mathcal{G}(\mathcal{A})^i$ for all $1 \leq i \leq n$. Let C_k denote the lcs of $\{C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^1)}, \ldots, C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^n)}\}$. The characterization of subsumption for $\mathcal{EL}_\neg$ yields homomorphisms φ_i from $\mathcal{G}(C_k)$ into $\mathcal{G}(C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)})$ for all $1 \leq i \leq n$. Now it is easy to see that $\pi_i \circ \varphi_i$ yields a homomorphism from $\mathcal{G}(C_k)$ into $\mathcal{G}(\mathcal{A})^i$, $1 \leq i \leq n$, each mapping the root of $\mathcal{G}(C_k)$ onto a. Hence, $a \in_\mathcal{A} C_k$.

Assume C' with $\mathsf{depth}(C') \leq k$ and $a \in_\mathcal{A} C'$. By Theorem 25, there exist homomorphisms ψ_i from $\mathcal{G}(C')$ into $\mathcal{G}(\mathcal{A})^i$ for all $1 \leq i \leq n$, each mapping the root of $\mathcal{G}(C')$ onto a. Since $\mathsf{depth}(C') \leq k$, these homomorphisms immediately yield homomorphisms ψ'_i from $\mathcal{G}(C')$ into $\mathcal{G}(C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)})$ for all $1 \leq i \leq n$. Now the characterization of subsumption yields $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)} \sqsubseteq C'$ for all $1 \leq i \leq n$, and hence $C_k \sqsubseteq C'$. Thus, $C_k \equiv \mathsf{msc}_{k,\mathcal{A}}(a)$.

Analogously, in case starting from a, no cycle can be reached in $\mathcal{A}$, we conclude $\mathsf{lcs}(C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^1)}, \ldots, C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^n)}) \equiv \mathsf{msc}_\mathcal{A}(a)$. Otherwise, with the same argument as in the proof of Theorem 18, it follows that the msc does not exist. □

In Example 23, we obtain two atomic completions, namely $\mathcal{G}(\mathcal{A})^1$ with $\ell^1(b_2) = \{P\}$, and $\mathcal{G}(\mathcal{A})^2$ with $\ell^2(b_2) = \{\neg P\}$. Now Theorem 27 implies $\mathsf{msc}_\mathcal{A}(a) \equiv \mathsf{lcs}(C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^1)}, C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^2)})$, which is equivalent to

$$P \sqcap \exists r.(P \sqcap \exists r.\neg P) \sqcap \exists r.(P \sqcap \exists r.\exists r.\neg P).$$

The examples showing the exponential blow-up of the size of k-approximations and msc's in $\mathcal{EL}$ can easily be adapted to $\mathcal{EL}_{\neg}$. However, we only have a double exponential upper bound (though we strongly conjecture that the size can again single-exponentially be bounded): the size of each tree (and the corresponding concept descriptions) obtained from an atomic completion is at most exponential, and the size of the lcs of a sequence of $\mathcal{EL}_{\neg}$-concept descriptions can grow exponentially in the size of the input descriptions [3].

Moreover, by an algorithm computing the lcs of the concept descriptions obtained from the atomic completions, the k-approximation (the msc) can be computed in double exponential time.

Corollary 28 *Let $\mathcal{A}$ be an $\mathcal{EL}_{\neg}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$.*

- *The k-approximation of a always exists. It may be of size $|\mathcal{A}|^k$ and can be computed in double-exponential time.*
- *The msc of a exists iff $\mathcal{A}$ is inconsistent, or starting from a, no cycle can be reached in $\mathcal{A}$. If the msc exists, its size may grow exponentially in $|\mathcal{A}|$, and it can be computed in double-exponential time. The existence of the msc can be decided in polynomial time.*

5 Most Specific Concepts in $\mathcal{ALE}$

As already mentioned in the introduction, the characterization of instance relationships could not yet be extended from $\mathcal{EL}_{\neg}$ to $\mathcal{ALE}$. Since these structural characterizations were crucial for the algorithms computing the (k-approximation of the) msc in $\mathcal{EL}$ and $\mathcal{EL}_{\neg}$, no similar algorithms for $\mathcal{ALE}$ can be presented here. However, we show that

1. given that N_C and N_R are finite sets, the $\mathsf{msc}_{k,\mathcal{A}}(a)$ always exists and can effectively be computed (cf. Theorem 29);
2. the characterization of instance relationships in $\mathcal{EL}$ is also sound for $\mathcal{ALE}$ (cf. Lemma 30), which allows for approximating the k-approximation; and
3. we illustrate the main problems encountered in the structural characterization of instance relationships in $\mathcal{ALE}$ (cf. Example 31).

The first result is achieved by a rather generic argument. Given that the signature, i.e., the sets N_C and N_R, are fixed and finite, it is easy to see that also the set of $\mathcal{ALE}$-concept descriptions of depth $\leq k$ built using only names from $N_C \cup N_R$ is finite (up to equivalence) and can effectively be computed. Since the instance problem for $\mathcal{ALE}$ is known to be decidable [16], enumerating this set and retrieving the least concept description which has a as instance, obviously yields an algorithm computing $\mathsf{msc}_{k,\mathcal{A}}(a)$.

Theorem 29 *Let N_C and N_R be fixed and finite, and let $\mathcal{A}$ be an $\mathcal{ALE}$-ABox built over a set N_I of individuals and $N_C \cup N_R$. Then, for $k \in \mathbb{N}$ and $a \in \mathsf{Ind}(\mathcal{A})$, the k-approximation of a w.r.t. $\mathcal{A}$ always exists and can effectively be computed.*

Note that the above argument cannot be adapted to prove the existence of the msc for acyclic $\mathcal{ALE}$-ABoxes unless the size of the msc can be bounded appropriately. Finding such a bound remains an open problem.

The algorithm sketched above is obviously not applicable in real applications. Thus, in the remainder of this section, we focus on extending the improved algorithms obtained for $\mathcal{EL}$ and $\mathcal{EL}_\neg$ to $\mathcal{ALE}$.

5.1 Approximating the k-approximation in $\mathcal{ALE}$

We first have to extend the notions description graph and description tree from $\mathcal{EL}_\neg$ to $\mathcal{ALE}$: In order to cope with value restrictions occurring in $\mathcal{ALE}$-concept descriptions, we allow for two types of edges, namely those labeled with role names $r \in N_R$ (representing existential restrictions of the form $\exists r.C$) and those labeled with $\forall r$ (representing value restrictions of the form $\forall r.C$). Again, there is a 1–1 correspondence between $\mathcal{ALE}$-concept descriptions and $\mathcal{ALE}$-description trees, and an $\mathcal{ALE}$-ABox $\mathcal{A}$ is translated into an $\mathcal{ALE}$-description graph $\mathcal{G}(\mathcal{A})$ just as described for $\mathcal{EL}$-ABoxes. The notion of a homomorphism also extends to $\mathcal{ALE}$ in a natural way. A homomorphism φ from an $\mathcal{ALE}$-description tree $\mathcal{H} = (V_H, E_H, v_0, \ell_H)$ into an $\mathcal{ALE}$-description graph $\mathcal{G} = (V, E, \ell)$ is a mapping $\varphi : V_H \longrightarrow V$ satisfying the conditions (1) and (2) on homomorphisms between $\mathcal{EL}$-description trees and $\mathcal{EL}$-description graphs, and additionally (3) $\varphi(v)\forall r\varphi(w) \in E$ for all $v\forall rw \in E_H$.

We are now equipped to formalize soundness of the characterization of instance relationships for $\mathcal{ALE}$.

Lemma 30 *Let $\mathcal{A}$ be an $\mathcal{ALE}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$ an individual in $\mathcal{A}$, and C an $\mathcal{ALE}$-concept description. Further, let $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$ denote the $\mathcal{ALE}$-description graph of $\mathcal{A}$ and $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ the $\mathcal{ALE}$-description tree of C. If there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$, then $a \in_{\mathcal{A}} C$.*

Proof: If $\mathcal{A}$ is inconsistent, nothing has to be shown. Let $\mathcal{A}$ be a consistent $\mathcal{ALE}$-ABox and $\mathcal{I}$ a model of $\mathcal{A}$. Let $C_a = \bigsqcap_{a:D\in\mathcal{A}} D$ and $\mathcal{G}(C_a) = (V_a, E_a, a, \ell_a)$. Now, $\mathcal{I} \models \mathcal{A}$ implies $a^\mathcal{I} \in C_a^\mathcal{I}$. We show $a^\mathcal{I} \in C^\mathcal{I}$ by induction on $\mathsf{depth}(C)$:

$\mathsf{depth}(C) = 0$: Then $C = Q_1 \sqcap \ldots \sqcap Q_n$ with $Q_i \in N_C \cup \{\neg P \mid P \in N_C\} \cup \{\top, \bot\}$. We show $a^\mathcal{I} \in Q_i^\mathcal{I}$ for all $1 \leq i \leq n$.

For $Q_i = \top$ nothing has to be shown.

Assume $Q_i = \bot$. This would imply $C_a \equiv \bot$ in contradiction to $\mathcal{I} \models \mathcal{A}$.

Assume $Q_i \in N_C$ or $Q_i = \neg P$ for some $P \in N_C$. Then, $P \in \ell(a)$ or $\neg P \in \ell(a)$. By definition of $\mathcal{G}(C_a)$ and C_a, we get $C_a \sqsubseteq P$ or $C_a \sqsubseteq \neg P$, respectively, and hence $a^\mathcal{I} \in P^\mathcal{I}$ or $a^\mathcal{I} \in (\neg P)^\mathcal{I}$.

$\mathsf{depth}(C) > 0$: Then $C = Q_1 \sqcap \ldots \sqcap Q_n \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_m.C_m \sqcap \forall s_1.D_1 \sqcap \ldots \sqcap \forall s_k.D_k$ with $Q_i \in N_C \cup \{\neg P \mid P \in N_C\} \cup \{\top\bot\}$. We show $a^\mathcal{I} \in C'^\mathcal{I}$ for all conjuncts C' on the top-level of C.

For $C' = Q_i$, the claim follows as for $\mathsf{depth}(C) = 0$.

Let $C' = \forall s_j.D_j$ and $w \in V_C$ the $\forall s_j$-successor of w_0 with $\mathcal{G}(C)(w) = \mathcal{G}(D_j)$. Since φ is a homomorphism, there exists a $\forall s_j$-successor v of a in $\mathcal{G}(\mathcal{A})$ with $\varphi(w) = v$. By definition of $\mathcal{G}(\mathcal{A})$, this node v is the root of a subtree of $\mathcal{G}(C_a)$. In particular, $C_a \sqsubseteq \forall s_j.C_{\mathcal{G}(C_a)(v)}$. Thus, $a^{\mathcal{I}} \in C_a^{\mathcal{I}}$ implies $a^{\mathcal{I}} \in (\forall s_j.C_{\mathcal{G}(C_a)(v)})^{\mathcal{I}}$. Obviously, restricting φ to the nodes in $\mathcal{G}(C)(w)$ yields a homomorphism from $\mathcal{G}(C)(w)$ into $\mathcal{G}(C_a)(v)$. By Theorem 4, it follows $C_{\mathcal{G}(C_a)(v)} \sqsubseteq C_{\mathcal{G}(C)(w)}$, and hence $a^{\mathcal{I}} \in (\forall s_j.D_j)^{\mathcal{I}}$.

Let $C' = \exists r_j.C_j$ and $w \in V_C$ the r_j-successor of w_0 with $\mathcal{G}(C)(w) = \mathcal{G}(C_j)$. Since φ is a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$, there exists an r_j-successor v of a in $\mathcal{G}(\mathcal{A})$ with $\varphi(w) = v$. If $v \notin \mathsf{Ind}(\mathcal{A})$, then y definition of $\mathcal{G}(\mathcal{A})$, this node v is the root of a subtree of $\mathcal{G}(C_a)$. As in the previous case, we get $a^{\mathcal{I}} \in (\exists r_j.C_j)^{\mathcal{I}}$. If $v \in \mathsf{Ind}(\mathcal{A})$, then φ restricted to the nodes in $\mathcal{G}(C)(w)$ yields a homomorphism ψ from $\mathcal{G}(C)(w)$ into $\mathcal{G}(\mathcal{A})$ with $\psi(w) = v$. By induction, we get $v \in_{\mathcal{A}} C_{\mathcal{G}(C)(w)}$. Since $\mathcal{I} \models \mathcal{A}$, it follows $v^{\mathcal{I}} \in C^{\mathcal{I}}_{\mathcal{G}(C)(w)}$, and since $C_j \equiv C_{\mathcal{G}(C)(w)}$, this yields $a^{\mathcal{I}} \in (\exists r_j.C_j)^{\mathcal{I}}$. □

As an immediate consequence of this lemma, we get $a \in_{\mathcal{A}} C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ for all $k \geq 0$, where the trees $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ and $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ are defined just as for $\mathcal{EL}$. This in turn yields $\mathsf{msc}_{k,\mathcal{A}}(a) \sqsubseteq C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ and hence, an algorithm computing an approximation of the k-approximation for $\mathcal{ALE}$. In fact, such approximations already turned out to be quite usable in our process engineering application [4].

The following example now shows that the characterization is not complete for $\mathcal{ALE}$, and that, in general, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))} \not\equiv \mathsf{msc}_{k,\mathcal{A}}(a)$. In particular, it demonstrates the difficulties one encounters in the presence of value restrictions.

Example 31 *Consider the $\mathcal{ALE}$-ABox*

$$\begin{aligned}\mathcal{A} \;:=\; & \{a : P, b_1 : P \sqcap \forall s.P \sqcap \exists r.P, b_2 : P \sqcap \exists r.(P \sqcap \exists s : P),\\ & (a, b_1) : r, (a, b_2) : r, (b_1, b_2) : r\},\end{aligned}$$

and the $\mathcal{ALE}$-concept description $C = \exists r.(\forall s.P \sqcap \exists r.\exists s.\top)$; $\mathcal{G}(\mathcal{A})$ and $\mathcal{G}(C)$ are depicted in Figure 5. Note that $\mathcal{G}(\mathcal{A})$ is the unique atomic completion of itself (w.r.t. $N_C = \{P\}$).

It is easy to see that there does not exist a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = a$. However, $a \in_{\mathcal{A}} C$: For each model $\mathcal{I}$ of $\mathcal{A}$, $b_2^{\mathcal{I}}$ does not have an s-successor, or at least one s-successor. In the first case, $b_2^{\mathcal{I}} \in \forall s.P$, and hence $b_2^{\mathcal{I}}$ yields the desired r-successor of $a^{\mathcal{I}}$ in $(\forall s.P \sqcap \exists r.\exists s.\top)^{\mathcal{I}}$. In the second case, it is $b_2^{\mathcal{I}} \in (\exists s.\top)^{\mathcal{I}}$, and hence $b_1^{\mathcal{I}}$ yields the desired r-successor of $a^{\mathcal{I}}$. Thus, for each model $\mathcal{I}$ of $\mathcal{A}$, $a^{\mathcal{I}} \in C^{\mathcal{I}}$.

Moreover, for $k = 4$, $C_{\mathcal{T}_4(a,\mathcal{A})}$ is given by $P \sqcap \exists r.(P \sqcap \forall s.P \sqcap \exists r.P \sqcap \exists r.(P \sqcap \exists r.(P \sqcap \exists s.P))) \sqcap \exists r.(P \sqcap \exists r.(P \sqcap \exists s.P))$. It is easy to see that $C_{\mathcal{T}_4(a,\mathcal{A})} \not\sqsubseteq C$. Hence, $C_{\mathcal{T}_4(a,\mathcal{A})} \sqcap C \sqsubset C_{\mathcal{T}_4(a,\mathcal{A})}$, which implies $\mathsf{msc}_{4,\mathcal{A}}(a) \sqsubset C_{\mathcal{T}_4(a,\mathcal{A})}$.

Intuitively, the above example suggests that, in the definition of atomic completions, one should take into account not only (negated) concept names

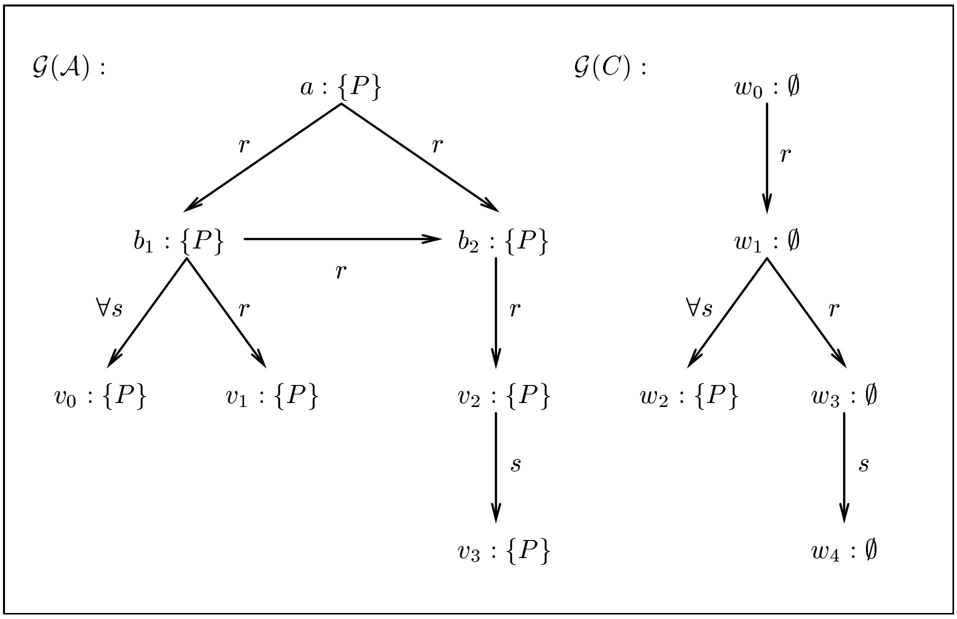

Figure 5: The $\mathcal{ALE}$-description graph and the $\mathcal{ALE}$-description tree from Example 31.

but also more complex concept descriptions. However, it is not clear whether an appropriate set of such concept descriptions can be obtained just from the ABox and how these concept descriptions need to be integrated in the completion in order to obtain a sound and complete structural characterization of instance relationships in $\mathcal{ALE}$.

6 Conclusion

Starting with the formal definition of the k-approximation of msc we showed that, for $\mathcal{ALE}$ and a finite signature (N_C, N_R), the k-approximation of the msc of an individual b always exists and can effectively be computed. For the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$, we gave sound and complete characterizations of instance relationships that lead to practical algorithms. As a by-product, we obtained a characterization of the existence of the msc in $\mathcal{EL}$-/$\mathcal{EL}_\neg$-ABoxes, and showed that the msc can effectively be computed in case it exists.

First experiments with manually computed approximations of the msc in the process engineering application were quite encouraging [4]: used as inputs for the lcs operation, i.e., the second step in the bottom-up construction of the knowledge base, they lead to descriptions of building blocks the engineers could use to refine their knowledge base. In next steps, the run-time behavior and the quality of the output of the algorithms presented here is to be evaluated by a prototype implementation in the process engineering application.

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases.* Addison-Wesley, 1995.

[2] F. Baader and R. Küsters. Computing the least common subsumer and the most specific concept in the presence of cyclic $\mathcal{ALN}$-concept descriptions. In *Proc. of KI'98*, LNAI 1504. Springer-Verlag, 1998.

[3] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In *Proc. of IJCAI'99.* Morgan Kaufmann, 1999.

[4] F. Baader and R. Molitor. Building and structuring description logic knowledge bases using least common subsumers and concept analysis. In *Proc. of ICCS2000*, LNAI 1867. Springer-Verlag, 2000.

[5] W.W. Cohen, A. Borgida, and H. Hirsh. Computing least common subsumers in description logics. In *Proc. of AAAI'92.* MIT Press, 1992.

[6] M. Chein and M. Mugnier. Conceptual graphs: Fundamental notions. *Revue d'Intelligence.* 6(4), 1992.

[7] W.W. Cohen and H. Hirsh. Learning the CLASSIC description logic: Theoretical and experimental results. In *Proc. of KR'94.* Morgan Kaufmann, 1994.

[8] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, 1979.

[9] V. Haarslev and R. Möller. Expressive ABox reasoning with number restrictions, role hierarchies and transitively closed roles. In *Proc. of KR2000.* Morgan Kaufmann, 2000.

[10] I. Horrocks. Using an expressive description Logic: FaCT or fiction? In *Proc. of KR'98.* Morgan Kaufmann, 1998.

[11] R. Küsters. *Non-Standard Inference in Description Logics.* PhD thesis, RWTH Aachen, 2000.

[12] R. Küsters. *Characterizing the Semantics of Terminological Cycles in $\mathcal{ALN}$ using Finite Automata.* In *Proc. of KR'98.* Morgan Kaufmann, 1998.

[13] R. Küsters and A. Borgida. What's in an attribute? Consequences for the least common subsumer. JAIR 2001. To appear.

[14] R. Küsters and R. Molitor. Computing least common subsumers in $\mathcal{ALEN}$. In *Proc. of IJCAI'01.* Morgan Kaufmann, 2001. To appear.

[15] B. Nebel. Reasoning and Revision in Hybrid Representation Systems. LNAI 422, Springer-Verlag, 1990.

[16] A. Schaerf. On the complexity of the instance checking problem in concept languages with existential quantification. *Journal of Intelligent Information Systems*, 2, 1993.

[17] L. von Wedel and W. Marquardt. ROME: A repository to support the integration of models over the lifecycle of model-based engineering processes. In *Proc. of ESCAPE-10*, 2000.